

Refine Predictions Ad Infinitum?

Mukund Sundararajan* Inbal Talgam-Cohen†

February 28, 2013

Abstract

We study how standard auction objectives in sponsored search markets change with refinements in the prediction of the relevance (click-through rates) of ads. We study mechanisms that optimize for a convex combination of efficiency and revenue. We show that the objective function of such a mechanism can only improve with refined (improved) relevance predictions, i.e., the search engine has no disincentive to perform these refinements. More interestingly, we show that under assumptions, refinements to relevance predictions can only improve the efficiency of any such mechanism. Our main technical contribution is to study how relevance refinements affect the similarity between ranking by virtual-value (revenue ranking) and ranking by value (efficiency ranking). Finally, we discuss implications of our results to the literature on signaling.

1 Introduction

Sponsored search is a multi-billion dollar market; it enables contextual advertising, and generates revenue that supports innovation in search algorithms. Besides being important, sponsored search markets are also technically interesting and have been investigated theoretically from several perspectives. For instance, auction theory (cf. [1, 7]), game theory (cf. [26, 6]), and bipartite matching theory (cf. [20]). See [16] for a survey.

How do these markets operate? Market efficiency (or value maximization) is achieved by displaying relevant ads that maximize the odds of the user clicking on the impression (a shown ad), and then successfully transacting on the advertiser’s website. To do this, the search engine must acquire two very different types of information. First, it must *estimate* the *relevance* of an advertiser to the user’s query, modeled as the probability that that advertiser’s ad will receive a click when it is shown to the user. Second, it must *elicit* in an incentive compatible way the *value* that the advertiser has for the user’s click; this quantity is determined usually by the the probability of transaction given a visit to the advertiser’s site, and the profit per transaction; notice that the search engine is not privy to either quantity. The *realized value* (value-per-impression) in this market is naturally modeled as a *product* of the value-per-click and relevance. Indeed, it is possible that an advertiser with a low value-per-click (v_1) and high probability of click (p_1) would realize a higher realized value than one with a high value-per-click (v_2) but low probability of click (p_2), because $v_1 p_1 > v_2 p_2$.

*Google Inc., Mountain View, CA 94043 USA.

†Computer Science Department, Stanford University, Stanford, CA 94305 USA. The authors wish to thank Amir Najmi for suggesting the problem, Mohammad Mahdian for suggesting that Pareto optimal mechanisms are virtual value-based, and Qiqi Yan for many helpful comments.

The elicitation problem mentioned above is naturally modeled via auction theory (cf. [1, 7]). The goal here is to maximize an auction objective such as efficiency or revenue by eliciting the value-per-click in an incentive compatible way. The estimation problem, however, is most naturally a machine learning problem [12], that is the goal is to *predict* relevance using features of the ad and the query. The relevance predictions are *refined* by either improving the machine learning algorithm or by adding new features. Consider an example of a refinement: Two pizza merchants, one from San Francisco and the other from the nearby city of San Jose, may appear equally relevant (both have say probability p) for the query ‘pizza’ emanating from an unspecified location in the Bay Area, but may have antisymmetric relevances on either side of p once the user’s location within this region is further pinpointed.

The focus of this paper is to study how standard auction objectives, specifically efficiency, behave with relevance prediction refinements like the one above.¹

Conventional wisdom would suggest that refinement ought to have a positive impact on the objective for which the auction is optimal. After all, why should more information hurt? Consistent with this intuition, it has been shown very generally that refinements can only improve the efficiency of the optimally efficient mechanism, or the revenue of the revenue-optimal mechanism [10].²

Things begin to get more interesting when we study changes in the revenue of the optimally efficient mechanism, or the efficiency of the revenue-optimal mechanism due to refinements (indeed, the market maker may wish to optimize any combination of revenue and efficiency - see discussion in Section 2.4). For instance, the revenue of the optimally efficient mechanism can fall with refinements. It is easiest to see this in a single-slot context with two bidders (like our ‘pizza’ example). Recall that the efficient auction allocates the slot to the advertiser with the highest realized value, and charges it the second highest realized value. The refinement suggested in the pizza example above causes relevances to become antisymmetric, and second highest realized value to drop, thereby reducing revenue.

Similarly, we can demonstrate that the efficiency of the revenue-optimal mechanism can fall with refinement – see Examples 2.2 and 3.2. What drives these examples? Recall (or consult Section 2.3) that the revenue-optimal auction and the efficient auction both rank bidders (ads) by a monotone function of their bids, ignoring ones for which this function is negative, and allocate the remaining bidders in this sequence to the available ad slots. The key difference is that the two mechanisms employ different functions of the bids. In these bad examples, with refinement, the revenue-optimal ranking drifts further apart from the efficient ranking, demonstrating that the twin objectives of revenue and efficiency are not necessarily aligned in the context of refinement.

Our first, comparatively straightforward result (Section 4) shows that once the search engine commits to a specific trade-off between efficiency and revenue, the resulting *Pareto optimal* mechanism can only benefit from refinement. Thus, the search engine has no disincentive to perform refinement.³

Our second, more technically challenging result (Section 3) is to identify assumptions

¹We study incentive-compatible mechanisms. The mechanisms used in practice, though not incentive compatible, have equilibria that are allocation- and revenue-equivalent to the corresponding incentive compatible mechanisms [6, 7]. So we expect our results to apply to practically used mechanisms *in equilibrium*.

²This paper shows that relevance prediction problem is decoupled from the value-per-click elicitation problem. That is, improvements to prediction improve the revenue of the natural Myerson mechanism.

³Obviously, this should not be at the cost of using features that violate user privacy.

under which refinement improves the *efficiency* of *every* Pareto optimal mechanism. The first assumption is that the value-per-click distributions are i.i.d. and satisfy the monotone hazard rate assumption, a fairly standard assumption. The second assumption is that that refinement causes the relevances of every pair of ads for every query to either reorder or grow further apart; this assumption is arguably restrictive and we discuss it in Section 3.2.5. We demonstrate that both assumptions are necessary via examples. Our main proof technique is to show that allocation ranking of every Pareto optimal mechanism draws closer to the efficient ranking as refinements are performed under the assumptions mentioned earlier.

Finally, our results are also applicable to the literature on signaling (cf. [21, 3, 8, 22]). Broadly, the connection can be described as follows: Realized value is a function of relevance, which can be modeled as the seller’s (search engine’s) private signal, and of the value-per-click, which is the bidder’s (advertiser’s) private signal. In our model, the seller knows that the bidder’s realized value changes multiplicatively (scales) with the seller’s signal. Therefore, he can elicit the bidder’s signal, perform the transformation and find the bidder’s realized value himself. Thus, the standard question in the signaling literature – how much of his signal should the seller reveal to the bidders – becomes a question of how refining predictions affects auction objectives. Our conclusion rephrased in the context of signaling is that, as long as realized values change multiplicatively with the signal and the above assumptions hold, revealing information improves both efficiency and revenue. See Section 5 for more detail.

2 The Model

2.1 The Sponsored Search Market

Position auctions, also known as slot auctions, keyword auctions or sponsored search auctions, are used for selling online advertisement slots that appear next to search results [1, 26, 6]. We use a standard model for position auctions [16], and modify it slightly to allow us to discuss the effect of refining relevance predictions.

2.1.1 Position Auctions Model

In the standard static model for position auctions, the single-dimensional private value of every advertiser is the amount he’s willing to pay per click on his ad. Advertisers’ values are transformed into values per impression, by multiplying them by click probabilities, known as click-through rates. We assume that the click-through rates are separable [16], i.e., can be written as a product of the advertiser’s relevance to the query, and the effect of the slot’s position on the page. Both components of the click-through rates are predicted by a machine learning system (see, e.g., [12]).

We use the following notation. For a search query q , the seller has m ad slots to sell to n advertisers. Advertiser i has a private value $v_i \in \mathbb{R}_+$, and a non-private click-through rate $p_{(q,i)}s_j$ for slot j , where $0 < p_{(q,i)} \leq 1$ is the query-advertiser relevance,⁴ and $1 \geq s_1 \geq \dots \geq s_m \geq 0$ are the decreasing slot effects. Where the query q is clear from the context we use p_i to denote the relevance.

⁴Our results hold when $p_{(q,i)}$ is allowed to have a zero value as well; to simplify the exposition we assume it is strictly positive.

The value per impression of advertiser i for slot j is $v_{i,j} = p_i s_j v_i$. The search engine ranks advertisers according to the component of $v_{i,j}$ that is not slot related:

Definition 2.1 (Realized value). *Advertiser i 's realized value for a slot in query q is*

$$r_{(q,i)} = p_{(q,i)} v_i$$

An important feature of the model is that the relation between the values and realized values is multiplicative.

Note that position auctions generalize multi-item auctions, in which one or more units of a single item are sold to unit-demand bidders: by setting $p_i = s_j = 1$ for every i, j we get a standard multi-item auction with m identical units.

2.1.2 A Model of Relevance Prediction

As described above, the relevance of each advertiser to a search query is predicted via a machine learning system. We now provide a model of the system's output, which we refer to as a *prediction scheme*.

The machine learning system has access to a collection of search query and advertiser *features*. Possible features include search keywords, geographic location, time, user data and search history, as well as ad text and landing page [12]. Adopting the standard assumption that features are discretized, consider the set of all possible query-advertiser pairs. These pairs are partitioned into types or *parts*, and for each part, the machine learning system produces an estimate of the (slot-independent) relevance probability p .⁵

For example, the following description defines a simple part: “pizza ad, user located in the Bay Area”. By taking into account additional features, coarse parts can be divided into finer *subparts*, such as “pizza ad, user located in San Francisco”. This process is called *refining* the partition.

We denote a partition by T and a part by t , and will often use the convention that \bar{T} is a coarse partition and \bar{t} a coarse part, whereas T is a refined partition and t is a subpart. Given any coarse part \bar{t} , there is a distribution over its subparts $\{t \mid t \subseteq \bar{t}\}$ arising from the underlying distributions of the features.

We can now define a prediction scheme, which is a partition and corresponding relevance predictions learned by the machine learning system:

Definition 2.2 (Prediction scheme). *A prediction scheme is a partition T of all query-advertiser pairs, and for every $t \in T$ a relevance prediction p_t for query-advertiser pairs in part t .⁶*

Overloading notation we denote both a prediction scheme and its partition by T .

How is a prediction scheme T applied in an auction for a query q ? For every query-advertiser pair (q, i) , we say that its relevance prediction $p_{(q,i)}$ is *according to T* if $p_{(q,i)} = p_t$, where $t \in T$ is the part to which the pair (q, i) belongs. The prediction scheme T is applied in an auction by setting the advertisers' relevance predictions according to T , finding the realized values and running the auction on these values.

⁵We also assume that this system produces the slot-specific relevance parameters (s_j 's), but we don't need to specify how this is done because our refinements apply only to the slot-independent terms (p_i 's).

⁶As defined, a prediction scheme is a deterministic clustering scheme [11]. In general a prediction scheme can also be randomized, containing a distribution over relevance predictions for every part [8, 22]. Our results hold for randomized prediction schemes as well.

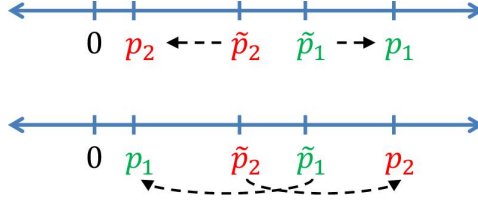


Figure 1: An example of spread (above) or flipped (below) relevance pairs (see Definition 2.4).

2.2 Prediction Refinements

An important auction design issue is how finer partitions of advertiser-query pairs may affect auction objectives. As we shall show, this affects the order in which the search engine ranks advertisers who are competing for slots.⁷ This in turn could have a significant effect on the revenue and efficiency of the auction. To study this effect we need a formal definition of prediction refinement.

Definition 2.3 (Refinement). *A prediction scheme T is a refinement of \bar{T} if its partition is a refinement of \bar{T} 's partition, and the average relevance of all subparts of a coarse part \bar{t} is its relevance according to \bar{T} , i.e.,*

$$p_{\bar{t}} = \mathbb{E}_{t \subseteq \bar{t}}[p_t].$$

If the subpart and its coarse counterpart are clear from context, we use p and \bar{p} to denote the corresponding relevance predictions.

We now define a natural class of refinements – those which distinguish among the advertisers, thus enabling a better matching of advertisements to the query.

Definition 2.4 (Spread or flipped pairs). *A pair a, b is spread with respect to a pair c, d if*

$$\frac{a}{b} \geq \frac{c}{d} \geq 1 \quad \text{or} \quad 1 \geq \frac{c}{d} \geq \frac{a}{b};$$

a, b is flipped with respect to c, d if

$$\frac{a}{b} \geq 1 \geq \frac{c}{d} \quad \text{or} \quad \frac{c}{d} \geq 1 \geq \frac{a}{b}.$$

See Figure 2.2 for an example.

Definition 2.5 (Flip-spread refinement). *A prediction scheme T is a flip-spread refinement of \bar{T} if T is a refinement of \bar{T} , and for every query q and pair of advertisers denoted without loss of generality 1 and 2, their relevance pair $p_{(q,1)}, p_{(q,2)}$ according to T is spread or flipped with respect to their relevance pair $\bar{p}_{(q,1)}, \bar{p}_{(q,2)}$ according to \bar{T} .*

In particular, *any* refinement is flip-spread for a prediction scheme \bar{T} that does not distinguish among advertisers: If all advertisers competing for a query q belong to the same part in \bar{T} and so appear equally relevant, then for every two advertisers it holds that $\bar{p}_1/\bar{p}_2 = 1$, and so any relevance pair is spread or flipped with respect to \bar{p}_1, \bar{p}_2 .

⁷We assume that the choice of T does not affect the *bias* of the predictions. Though very fine parts could cause inaccuracies due to lack of data.

Example 2.1. [Flip-spread refinement] Consider the position auction described in the introduction, in which two pizzerias – the first located in San Francisco (SF) and the second in San Jose (SJ) – compete for a single advertisement slot next to search results for query ‘pizza’. Let \bar{T} be a coarse prediction scheme. Assume that both pizzerias are equally popular, and so both query-advertiser pairs belong to the same part \bar{t} in \bar{T} (described as “pizza ad, user located in the Bay Area”), and their associated relevances are $\bar{p}_1 = \bar{p}_2 = \bar{p}_{\bar{t}} = 0.75$. As mentioned above, this makes any refinement of \bar{T} flip-spread.

Now assume the search engine has access to a more precise location feature of the query q , indicating whether the user is in SF ($q = \text{SF}$) or in SJ ($q = \text{SJ}$), and each occurs with equal probability $1/2$. When the prediction scheme is refined by including this feature, the relevances p_1, p_2 according to the refined scheme T behave antisymmetrically: If the query comes from a user in SF, the relevance $p_{(\text{SF},1)}$ of the SF advertiser is 1 and the relevance $p_{(\text{SF},2)}$ of the SJ advertiser is 0.5. If the query comes from a user in SJ, $p_{(\text{SJ},1)} = 0.5$ and $p_{(\text{SJ},2)} = 1$. Indeed, in both cases the pair p_1, p_2 is either spread or flipped with respect to \bar{p}_1, \bar{p}_2 .

Example 2.2. [Non-flip-spread refinement] Consider again a single-slot position auction for ‘pizza’. Assume now that advertiser 1 is a nationwide chain of pizzerias while advertiser 2 is a local artisan pizzeria in SF. Consider a coarse prediction scheme \bar{T} as above, and a refinement T where this time the refining feature indicates whether $q = \text{SF}$ (happens with probability $(1/4) - \delta$ for some small δ) or $q = \neg \text{SF}$ (happens with probability $(3/4) + \delta$). The relevance of advertiser 1 does not depend on user location and his relevance predictions are $\bar{p}_1 = p_{(\text{SF},1)} = p_{(\neg \text{SF},1)} = 0.8$. On the other hand, the relevance predictions for advertiser 2 are $\bar{p}_2 = 0.1$, which is the average of $p_{\text{SF},2} = 0.4$ and $p_{\neg \text{SF},2} = \epsilon$ (this is by setting $\delta = 15\epsilon/(8 - 20\epsilon)$). Refinement T is not flip-spread, since for $q = \text{SF}$, the advertisers relevance predictions 0.8, 0.4 are not spread or flipped with respect to the coarser predictions 0.8, 0.1.

2.3 Virtual Value Based Mechanisms

2.3.1 The Bayesian Setting

We assume a Bayesian setting in which the advertisers’ values are i.i.d., and are drawn from a publicly known prior distribution F with a positive smooth density f . The advertisers are not symmetric since they may have different click-through rates and thus non-i.i.d. realized values.

Given F with density f from which value v_i is drawn, the *inverse hazard rate* (or *information rent*) of v_i is

$$\lambda^F(v_i) = \frac{1 - F(v_i)}{f(v_i)}.$$

The *virtual value* corresponding to v_i is $\varphi^F(v_i) = v_i - \lambda^F(v_i)$. A similar definition applies to the realized value: Given a distribution G with density g from which the realized value r_i is drawn, the *realized virtual value* of r_i is $\varphi^G(r_i) = r_i - \lambda^G(r_i)$. The relation $r_i = p_i v_i$ between the value and realized value implies that

$$G(r_i) = F(v_i), \quad g(r_i) = \frac{1}{p_i} f(v_i), \quad \varphi^G(r_i) = p_i \varphi^F(v_i).$$

Note that MHR (regular) values imply MHR (regular) realized values. We omit F from the

notation φ^F, λ^F where the distribution is clear from the context,⁸ and where v_i is clear from the context we may use the notation $\varphi_i = \varphi(v_i)$.

2.3.2 Standard Assumptions on the Prior Distribution

A distribution F is *MHR* (*montone hazard rate*) if its inverse hazard rate function $\lambda(\cdot)$ is non-increasing, and is *regular* if its virtual value function $\varphi(\cdot)$ is non-decreasing. We say that values are MHR (regular) if they're drawn from an MHR (regular) distribution, and that a position auction is MHR (regular) if its advertisers' values are MHR (regular).

The assumption of MHR values is standard in the mechanism design literature (see, e.g., [19]). Many commonly studied distributions are MHR, including the uniform, exponential and normal distributions, and all distributions with log-concave densities [9]. Every MHR distribution is regular but not vice versa; an example of a regular but non-MHR distribution is the *equal revenue distribution*, defined as $F(v) = 1 - \frac{1}{v}$.

2.3.3 Welfare and Revenue-Optimal Mechanisms

Mechanisms are evaluated according to their performance in terms of expected welfare and expected revenue, where the expectation is over the value distribution and the query distribution. We discuss briefly the form of the efficiency-optimal and revenue-optimal auctions for sponsored search as identified by prior literature (cf. [1, 7]).

The *VCG* auction [27, 4, 13] maximizes expected social welfare among all truthful and individually rational (IR) mechanisms by finding the most efficient allocation. Every bidder is charged his externality – the difference between the maximum welfare if he does not participate in the auction and the welfare of all other bidders when he does. In the context of position auctions, the most efficient allocation assigns the m advertisers with highest realized values r_i to the m slots, after ordering them from high to low (we assume throughout that ties are broken lexicographically). See [1] for the exact form of the VCG prices in the sponsored search setting.

The *revenue-optimal* mechanism [23] maximizes expected revenue among all truthful and IR mechanisms. To find the optimal allocation, Myerson proved the following key lemma for single-item auctions, which relates the expected revenue of an allocation rule to the (realized) *virtual surplus* served by it. Recall that an allocation rule is *monotone* if for every bidder i and every fixed set of bids by bidders other than i , bidder i 's expected allocation (weakly) increases with its bid. We now restate Myerson's lemma in the context of position auctions.

Lemma 2.6 (Myerson [23] for sponsored search auctions). *Every truthful, IR position auction has a monotone allocation rule. Moreover, its expected revenue is equal to its expected realized virtual surplus, i.e., $\mathbb{E}_{\mathbf{v}}[\sum_{i,j} s_j \varphi_i(r_i) x_{i,j}(\mathbf{v})]$ where $x_{i,j}(\mathbf{v})$ indicates if bidder i wins slot j given value profile \mathbf{v} .*⁹

When values are regular, the Myerson mechanism maximizes expected revenue by assigning up to m slots to the $\leq m$ advertisers with highest *non-negative* realized virtual values, in high

⁸In particular, by $\varphi(v_i)$ we refer to the virtual value and by $\varphi(r_i)$ we refer to the realized virtual value.

⁹Myerson also showed that there is a unique pricing rule that yields a truthful IR mechanism when coupled with a monotone allocation rule. Notice however that the form of this pricing rule is rendered unimportant because this lemma gives us a handle on revenue even without knowing the precise form of the prices.

to low order.¹⁰ By regularity, if an advertiser's value increases then his position can only increase, and so the allocation rule is monotone, hence truthful.

2.3.4 Definition of Virtual Value Based Mechanisms

We now define a class of truthful mechanisms of which Myerson and VCG are extremal members. In Section 2.4 we will show that these mechanisms optimize convex combinations of efficiency and revenue.

Let v be an advertiser's value drawn from the distribution F . Recall that $\lambda(v)$ is the inverse hazard rate of v .

Definition 2.7 (α -virtual value). *For $\alpha \geq 0$, the α -virtual value of v is*

$$\varphi^\alpha(v) = v - \alpha\lambda(v).$$

Observe that the α -virtual value can be rewritten as a combination of the value and virtual value: $\varphi^\alpha(v) = (1 - \alpha)v + \alpha v - \alpha\lambda(v) = (1 - \alpha)v + \alpha\varphi(v)$. For $\alpha = 0$ we get the value and for $\alpha = 1$ we get the virtual value.

Definition 2.8 (α -virtual value based mechanism). *For $\alpha \geq 0$, the α -virtual value based mechanism asks the advertisers to report their values v_i , ranks them according to the realized α -virtual values $p_i\varphi_i^\alpha$, and allocates the slots to the advertisers with highest non-negative such values.*

The resulting mechanism is deterministic. For $\alpha = 0$ we get the VCG auction, and for $\alpha = 1$ and regular values we get the Myerson mechanism.

Lemma 2.9 (Truthfulness of α -virtual value based mechanism). *For $0 \leq \alpha \leq 1$ and regular values, the α -virtual value based mechanism is truthful.*

Proof. Since $\varphi^\alpha = (1 - \alpha)v + \alpha\varphi(v)$ and the value distribution is regular, φ^α is non-decreasing in v when $0 \leq \alpha \leq 1$, and so the allocation rule of the α -virtual value based mechanism is monotone; truthfulness follows from Lemma 2.6. \square

2.4 Pareto Optimal Mechanisms

In this section we consider mechanisms that are optimal with respect to a combination of expected welfare and revenue. Such mechanisms are termed Pareto optimal, as they lie on the Pareto frontier of welfare and revenue. Myerson and Satterthwaite [24] originally established the form of Pareto optimal mechanisms in a slightly different context; they used such mechanisms to guarantee individual rationality and incentive compatibility for bilateral trade. We study these mechanisms because we expect the search engine to use a mechanism from this family. Naively, you would think that the search engine ought only to care about its own revenue, but revenue as modeled in this paper is just 'short-term' revenue. In the long run, search engines are invested in the health of their ad markets, and it is reasonable to assume

¹⁰Lahaie and Pennock [15] study methods for optimizing revenue of sponsored search auctions given a practically motivated constraint, that we rank by functions of the form $p_i^\alpha v_i$. They note that Myerson's virtual value based approach does not fit this scheme.

that they care about more than just short-term revenue, i.e., for instance, a combination of efficiency and revenue.¹¹

We now discuss the form of Pareto optimal mechanisms for sponsored search auctions and show that they are virtual value based. That is, for regular value distributions and for every fixed trade-off $(1 - \alpha)\mathbb{E}[\text{welfare}] + \alpha\mathbb{E}[\text{revenue}]$, the α -virtual value based mechanism is optimal among all truthful mechanisms.¹² Thus, every point on the Pareto frontier among welfare and revenue can be realized by a virtual value based auction.

Recall the standard *rearrangement inequality*:

Lemma 2.10. *For every two decreasing vectors \mathbf{r} and \mathbf{s} such that $r_1 \geq \dots \geq r_n$ and $s_1 \geq \dots \geq s_n$, and for every ranking (permutation) π of $\{1, \dots, n\}$,*

$$\sum_i s_i r_{\pi_i} \leq \sum_i s_i r_i. \quad (1)$$

Lemma 2.11 (Pareto optimal mechanism). *Consider a regular position auction and a convex combination objective $(1 - \alpha)\mathbb{E}[\text{welfare}] + \alpha\mathbb{E}[\text{revenue}]$, where $0 \leq \alpha \leq 1$. Then the optimal mechanism for this objective among all truthful and IR mechanisms is the α -virtual value based mechanism.*

Proof. Applying Myerson's lemma (Lemma 2.6) we get that the mechanism's objective is to maximize

$$\begin{aligned} (1 - \alpha)\mathbb{E}_{\mathbf{v}}\left[\sum_{i,j} v_{i,j} x_{i,j}(\mathbf{v})\right] + \alpha\mathbb{E}_{\mathbf{v}}\left[\sum_{i,j} s_j \varphi(r_i) x_{i,j}(\mathbf{v})\right] = \\ \mathbb{E}_{\mathbf{v}}\left[\sum_{i,j} x_{i,j}(\mathbf{v}) \cdot ((1 - \alpha)s_j r_i + \alpha s_j \varphi(r_i))\right] \end{aligned}$$

Therefore, by the rearrangement inequality, the optimal allocation rule takes up to m bidders with highest non-negative combinations $(1 - \alpha)r_i + \alpha\varphi(r_i)$, and assigns them one by one to the highest slots. This is precisely the α -virtual value based mechanism, which is guaranteed to be truthful by Lemma 2.9. \square

3 How Does Refinement Affect Market Efficiency?

In this section we consider the effect of refined relevance prediction on the welfare guarantees of Pareto optimal mechanisms; see Section 2.4 for a discussion of why we think search engines would use a mechanism from this class. Thus, our goal is to study how the market is affected by refined relevance predictions. In our main technical result, we identify natural sufficient conditions under which refining the prediction improves the welfare of any Pareto optimal mechanism.

¹¹As Likhodedov and Sandholm [18] discuss in the context of multi-unit auctions, such mechanisms can be used to to maximize expected welfare subject to a minimum constraint on the expected revenue. More recently, Diakonikolas et al. [5] study the computational complexity of implementing such mechanisms for single-item auctions; they show that the problem is NP-hard, and provide an FPTAS for the two bidder case.

¹²This holds for irregular value distributions as well, where α -virtual value is replaced by ironed α -virtual value.

Theorem 3.1 (Refined prediction improves welfare). *Consider an i.i.d., MHR position auction and let M be a Pareto optimal mechanism. Then for every value profile \mathbf{v} , coarse prediction scheme \bar{T} and flip-spread refinement T , running M with T increases welfare in comparison to running M with \bar{T} .*

Note that this result holds entirely *pointwise*, that is, it does not require averaging over the value profiles, nor does it require averaging over the subparts of T to which the query-advertiser pairs belong to (the latter is required by Theorem 4.1 on improving welfare-revenue trade-offs).

What is the technical content of this theorem? Recall that in a single-item auction setting when values are drawn i.i.d. from a regular (or MHR) distribution, the revenue-optimal and the efficient auction both rank bidders in the same order; of course, the revenue optimal mechanisms excludes bidders with negative virtual values by using a reserve price. In the sponsored search context, even when the value per click distributions are IID and mhr, the presence of the relevance terms cause the revenue-optimal ranking to differ from the optimal one. What we will show is that the difference between the revenue and efficiency rankings (or more precisely between the efficient and Pareto optimal rankings) diminishes with refinements so long as the conditions stated in the theorem statement above hold.

3.1 Proof of Main Result

In Pareto optimal auctions, bidders are ranked according to their α -virtual values $\varphi^\alpha = v - \alpha\lambda$. We can thus think of $\alpha * \lambda$ as a “penalty” on the value imposed by the seller. In expectation, the inverse hazard rate is the rent a winning bidder manages to keep to himself out of his total value when the seller applies the optimal mechanism. Thus a seller aiming to maximize expected revenue ‘penalizes’ bidders with large rents and demotes them. The monotonicity of the relative size of the penalty in comparison to the value plays an important role in our analysis; we will show that when this quantity is falling, refinements bring the revenue optimal ordering closer to the efficient one.

Definition 3.2 (Bid penalty fraction). *For every value $v \sim F$ with inverse hazard rate λ , its penalty fraction is λ/v .*

For MHR distributions, the penalty fraction is non-increasing: I.e., for every two values $v_1 > v_2$ drawn from an MHR distribution F ,

$$\frac{\lambda_1}{v_1} \leq \frac{\lambda_2}{v_2}, \quad (2)$$

since the inverse hazard rates satisfy $\lambda_1 \leq \lambda_2$.

However, monotonicity of the penalty fraction does not hold for all regular distributions – see distribution in Example 3.2. In a sense, mhr characterizes distributions with non-increasing penalty fractions: If the hazard rate was decreasing (i.e., the inverse hazard rates are increasing), it would be possible to shift the distribution far enough to the right so that the penalty fractions are increasing.

If, as for MHR distributions, the relative effect of the penalty diminishes as the value grows larger, then values grow further apart after transforming them into α -virtual values. Let v_1, v_2

be two i.i.d. MHR values, and let their α -virtual values be $\varphi_1^\alpha, \varphi_2^\alpha$. Then we claim that

$$\frac{v_1}{v_2} < \frac{\varphi_1^\alpha}{\varphi_2^\alpha} \implies v_1 > v_2. \quad (3)$$

We now prove our claim. Assume for contradiction that $v_1 < v_2$ and $v_1\varphi_2^\alpha < v_2\varphi_1^\alpha$. Plugging in $\varphi_i^\alpha = v_i - \alpha\lambda_i$ we get

$$v_1v_2 - \alpha v_1\lambda_2 < v_1v_2 - \alpha v_2\lambda_1 \iff v_1\lambda_2 > v_2\lambda_1,$$

achieving a contradiction by Inequality 2.

The next lemma is our key lemma and shows that the ranking of any Pareto optimal mechanism becomes more similar to the efficient ranking with refinement. That is, it shows that if the Pareto optimal mechanism's allocation is inefficient despite using a flip-spread refined relevance prediction, then its allocation necessarily remains inefficient when the prediction is not refined.

Lemma 3.3 (Inefficient allocation with refined prediction). *Let prediction scheme T be a flip-spread refinement of \bar{T} , and consider two advertisers 1 and 2 whose values are drawn from an IID regular distribution. Then*

$$p_1v_1 < p_2v_2 \text{ and } p_1\varphi_1^\alpha \geq p_2\varphi_2^\alpha > 0 \implies \bar{p}_1\varphi_1^\alpha \geq \bar{p}_2\varphi_2^\alpha.$$

In words, if advertiser 1 has lower realized value than 2 but higher positive realized α -virtual value according to the refined prediction T , then this holds according to \bar{T} as well.

Proof. First note that $v_2 > 0$, because advertiser 2's realized value is strictly higher than advertiser 1's. Combining the two inequalities on the lefthand side we get

$$\frac{\varphi_1^\alpha}{\varphi_2^\alpha} \geq \frac{p_2}{p_1} > \frac{v_1}{v_2}.$$

Because the value distribution is mhr, we can applying Equation 3 to show that $v_1 > v_2$, and so:

$$\frac{\varphi_1^\alpha}{\varphi_2^\alpha} \geq \frac{p_2}{p_1} > \frac{v_1}{v_2} > 1. \quad (4)$$

By definition of a flip-spread refinement, the pair p_1, p_2 is spread or flipped with respect to \bar{p}_1, \bar{p}_2 . So

$$\frac{p_2}{p_1} > 1 \implies \frac{p_2}{p_1} \geq \frac{\bar{p}_2}{\bar{p}_1}. \quad (5)$$

Equations 4 and 5 combined show that $\bar{p}_1\varphi_1^\alpha \geq \bar{p}_2\varphi_2^\alpha$, completing the proof. \square

We now state and prove a technical lemma that generalizes the classic rearrangement inequality (Lemma 2.10). Consider two orderings π_1, π_2 of the same ground set $(1, \dots, n)$. We say that π_1 is *more ordered* than π_2 if for every pair of elements $i > j$ which appear in order in π_2 , they also appear in order in π_1 , i.e., $\pi_2(i) > \pi_2(j) \implies \pi_1(i) > \pi_1(j)$.

Lemma 3.4 (Generalized rearrangement inequality). *Rename the advertisers such that their realized values are decreasing, i.e., $r_1 \geq \dots \geq r_n$. Let π_1, π_2 be two rankings of the advertisers such that π_1 is more ordered than π_2 . Let \mathbf{s} be a vector of n decreasing slot effects $s_1 \geq \dots \geq s_n$. Then $\mathbf{s} \cdot \mathbf{r}(\pi_1) > \mathbf{s} \cdot \mathbf{r}(\pi_2)$.*

For the proof we use the following notation: A ranking π is just an ordered vector of the advertisers $1, \dots, n$. We use $i < j$ for advertisers and $x < y$ for their ranks. So π_x is the advertiser that appears in rank x in π , and $\pi(i)$ is the rank of advertiser i in π . The notation $\mathbf{r}(\pi)$ means that we are ordering the vector \mathbf{r} in the same order in which the advertisers appear in π , that is, $r_{\pi_1}, \dots, r_{\pi_n}$. So

$$\mathbf{s} \cdot \mathbf{r}(\pi) = \sum_{x=1}^n s_x r_{\pi_x} = \sum_{i=1}^n s_{\pi(i)} r_i.$$

Proof. We first prove the statement for two rankings π^1, π^2 , which are identical except for two advertisers $i < j$ that appear consecutively in both but in flipped order. I.e, let $x = \pi^1(i)$ and $y = \pi^1(j)$, then $x = y - 1$ and

$$x = \pi^2(j) = \pi^2(i) - 1 = y - 1$$

(we are using here the assumption that π^1 is more ordered than π^2 and so i must appear before j in π^1).

Observe that for such rankings, since the only difference is advertisers $i < j$ appearing in consecutive ranks $x < y$ in π^1 and in flipped ranks y, x in π^2 , to show that $\mathbf{s} \cdot \mathbf{r}(\pi^1) \geq \mathbf{s} \cdot \mathbf{r}(\pi^2)$ it's sufficient to show

$$s_x r_i + s_y r_j \geq s_y r_i + s_x r_j. \quad (6)$$

Since vectors \mathbf{s}, \mathbf{r} are decreasing, $s_x \geq s_y$ and $r_i \geq r_j$, and so by the standard rearrangement inequality Condition 6 holds.

We now turn to general rankings π^1, π^2 . To show that $\mathbf{s} \cdot \mathbf{r}(\pi^1) \geq \mathbf{s} \cdot \mathbf{r}(\pi^2)$, we conceptually run a “bubble sort” on π^2 to turn it step by step into π^1 . In every step, we compare a pair of adjacent advertisers in π^2 and swap them if their order does not match that of π^1 . The proof is complete by noticing that since π^1 is more ordered to begin with, then π^2 is becoming increasingly more ordered along the process, because two advertisers are swapped only if they're in the wrong order. Therefore we can apply the above claim to get that $\mathbf{s} \cdot \mathbf{r}(\pi^2)$ is non-decreasing with each step, thus completing the proof. \square

We now show that in the more general setting of Theorem 3.1, inefficiencies due to refinement never occur.

of Theorem 3.1. By assumption, the n advertisers have i.i.d., MHR values. We can assume that $m \geq n$, i.e., that there are enough slots for all the ads. This is without loss of generality since the click-through rates of the lowest slots can be set to zero, by setting their slot effects $s_j = 0$. Observe that in a position auction with enough slots, an assignment of the advertisers to the slots is inefficient if and only if there are two advertisers who are assigned slots and for whom the following holds: the advertiser with the lower realized value among the two gets the higher slot, and vice versa.

Let the Pareto optimal mechanism M be the α -virtual value based mechanism which ranks the advertisers according to φ^α (Lemma 2.11). Fix a value profile \mathbf{v} , and for every advertiser, a query-advertiser part and subpart according to \bar{T} and its flip-spread refinement T . This fixes for every advertiser i the relevance predictions \bar{p}_i according to \bar{T} and p_i according to T .

Consider the assignment chosen by M when using the flip-spread refinement T for query-advertiser relevance predictions. Assuming the chosen assignment is inefficient, without loss of generality let advertisers 1 and 2 be such that both are assigned, and advertiser 1 has a lower realized value but higher slot than advertiser 2. Formally, $\varphi_1^\alpha, \varphi_2^\alpha \geq 0$, $p_1 v_1 < p_2 v_2$ and $p_1 \varphi_1^\alpha \geq p_2 \varphi_2^\alpha$.

If $\varphi_2^\alpha > 0$, we can now invoke Lemma 3.3 to get that $\bar{p}_1 \varphi_1^\alpha \geq \bar{p}_2 \varphi_2^\alpha$. If $\varphi_2^\alpha = 0$, the same inequality holds trivially. We have shown that if advertiser 1 is ranked above advertiser 2 by mechanism M using the flip-spread refinement T , then M will also rank 1 above 2 when using the coarse prediction scheme \bar{T} .

Let π_1 be the ranking of advertisers by M using T , and let π_2 be the ranking of advertisers by M using \bar{T} . We have shown that π_1 is more ordered than π_2 . We now apply Lemma 3.4. Note that the realized values are calculated with respect to the refined relevance prediction T , since these are the values that realize the welfare of M . Lemma 3.4 shows that running M with the flip-spread refinement T instead of \bar{T} increases the welfare, thus completing the proof. \square

3.2 Bad Examples and Necessity of Assumptions

We now demonstrate via examples that the assumptions like the ones made in Theorem 3.1 are necessary. All examples use two advertisers, one ad position, and use the revenue-optimal mechanism.

Recall the position auction described in Example 2.1. We assume both advertisers' values are i.i.d. and drawn from a regular distribution F with density f . The realized values of the advertisers are as follows:

1. $r(\text{SF},1) = v_1$, $r(\text{SJ},1) = v_1/2$;
2. $r(\text{SJ},2) = v_2$, $r(\text{SF},2) = v_2/2$;

For the analysis we fix two values $v \geq v' > 0$ drawn from F , but do not specify which value belongs to which advertiser (the advertisers are i.i.d. so each of the two possibilities occurs with probability $\frac{1}{2}$). Let φ, φ' be the virtual values corresponding to v, v' ; by regularity of F , $\varphi \geq \varphi'$. We assume that φ' is non-negative and φ is positive, since otherwise refinement does not change the allocation of the revenue-optimal mechanism, and therefore has no effect on welfare. We also assume without loss of generality that the user is from SJ (the other case is symmetric).

3.2.1 When refinement reduces efficiency

Assume $v > v'$ (if $v = v'$ then allocation using the user's location is always efficient). We show that a necessary condition for inefficiencies due to refinement to occur is

$$\frac{v'}{v} < \frac{1}{2} < \frac{\varphi'}{\varphi}, \quad (7)$$

and that if Condition 7 holds, then efficiency loss of magnitude $v/2 - v'$ occurs with probability $\frac{1}{2}$.

Assume first that Condition 7 holds. With probability $1/2$, the SF advertiser has the higher among the two values, and its realized value $v/2$ is higher than the other realized value v' by Condition 7. If the location data is ignored, the SF advertiser with the higher value wins and the assignment is efficient. But if the location is used for refinement, the SJ advertiser's virtual value φ' is higher than $\varphi/2$ (Condition 7), and so there is efficiency loss of $v/2 - v'$.

We now show that condition 7 is necessary, and there are no other cases of inefficient allocation due to refinement. Without refinement, the advertiser with the higher value wins, so an inefficiency due to refinement occurs only if the advertiser with the lower value wins and only if his realized value is lower. This yields Condition 7. Now consider the case in which the SJ advertiser has the higher value. The SF advertiser cannot win after refinement since its virtual value $\varphi'/2$ is lower than φ , so in this case there is no inefficiency due to refinement.

We summarize the total expected loss in welfare from inefficiencies due to refinement:

$$\int_{v_{\min}}^{v_{\max}} \int_{2v'}^{\bar{v}} (v/2 - v') f(v) f(v') dv dv', \quad (8)$$

where $[v_{\min}, v_{\max}]$ is the range of F , and \bar{v} is defined to be the value such that the corresponding virtual value $\varphi(\bar{v})$ is equal to $2\varphi'$.

3.2.2 When refinement increases efficiency

A similar analysis to the above shows that a necessary condition for refinement to increase efficiency is:

$$\frac{1}{2} < \min\left\{\frac{v'}{v}, \frac{\varphi'}{\varphi}\right\}. \quad (9)$$

When Condition 9 holds, then efficiency loss of magnitude $v/2 - v'$ occurs with probability $1/2$: If the SF advertiser gets the higher value, condition 9 implies that its realized value is *lower*. As above, with no refinement the SF advertiser wins and with refinement the SJ advertiser wins. So in this case lack of refinement leads to an efficiency loss of $v/2 - v'$.

The total expected loss in welfare from inefficiencies due to coarseness is

$$\int_{v_{\min}}^{v_{\max}} \int_{v_{\min}}^{2v'} (v/2 - v') f(v) f(v') dv dv'. \quad (10)$$

3.2.3 MHR example – refinement increases efficiency

Example 3.1 (MHR values). Let F be the uniform distribution over $[0, 1]$. There are no inefficiencies due to refinement since Condition 7 never holds:

$$\frac{v'}{v} < \frac{\varphi'}{\varphi} = \frac{2v' - 1}{2v - 1} \iff v' > v,$$

in contradiction to the definition of v, v' by which $v \geq v'$.

3.2.4 Non-MHR example – refinement can reduce efficiency

Example 3.2 (Non-MHR values). Let $H = 10^3$ be a truncating parameter and $b = -1$ be a shifting parameter. Let F be a variant of the equal revenue distribution achieved by truncating its support from $[1, \infty)$ to $[1, H]$, and shifting it to the right by $|b|$ (if $H = \infty$ and $b = 0$ we get the standard equal revenue distribution). The truncation ensures that F has finite expectation so that Myerson’s theory (Lemma 2.6) applies; the reason for shifting will become apparent below. The resulting distribution over the support $[2, H + 1]$ is

$$F(v) = \frac{H}{H-1} \left(1 - \frac{1}{v+b}\right).$$

F is regular since its virtual valuation function $\varphi(v) = ((v+b)^2/H) - b$ is increasing; note that since $b < 0$, $\varphi(v)$ is also strictly positive for every v . However, F is not MHR, since its inverse hazard rate function $\lambda(v) = v - ((v+b)^2/H) + b$ increases with v in the range $[2, 501]$.

By calculating and comparing the integrals in Equations 8 and 10, we show that the expected loss due to refinement is higher than that due to coarseness, and so in this example, the flip-spread refinement that takes into account user location hurts the expected welfare of the revenue-optimal mechanism. Details of the calculation appear in the appendix.

3.2.5 Non-Flip-Spread Refinements

We now consider the slightly modified setting of Example 2.2. Recall that in this setting, the user location sometimes made the advertisers seem more “similar”, and as a result, in more direct competition. To take advantage of the direct competition, the revenue-optimal mechanism sometimes allocates to the advertiser with lower realized value, increasing the expected revenue but decreasing the welfare, as demonstrated below.

First, we show that *pointwise* over the bids, the flip-spread assumption is necessary in a strong sense for a result such as Theorem 3.1. That is, if the ad relevances draw closer together without flipping, for every non-degenerate mhr distribution, there exist valuations for which efficiency of the revenue-optimal mechanism falls with refinement. To construct such an example, for any refined relevance pair $p_1 > p_2$, you could pick two values v_1 and $v_2 > v_1$ such that both virtual values are positive but near-zero, the realized value after refinement of the first advertiser ($p_1 \cdot v_1$) dominates the second ($p_2 \cdot v_2$), but the realized virtual value of the second advertiser ($p_2 \cdot \varphi_2$) slightly exceeds that of the first ($p_1 \cdot \varphi_1$). Thus after refinement, the revenue-optimal ranking disagrees with the efficient one. Now if the relevance pair p_1, p_2 is not flip-spread with respect to the unrefined relevance pair \bar{p}_1, \bar{p}_2 , the unrefined revenue ranking will agree with the efficient one, and thus refinement causes efficiency loss.

We now show for a specific selection of relevance parameters, refinements, and value distributions, that efficiency loss can happen *in expectation* when the flip spread assumption is violated. We use the relevance probabilities from Example 2.2. Assume that the advertisers’ values v_1, v_2 are drawn independently from the MHR uniform distribution over range $[3, 5]$. The ranges of their realized values and virtual values are as follows: Since his relevance prediction is 0.8 whether or not the prediction scheme is refined, advertiser 1’s realized value range is $[2.4, 4]$ and his virtual value range is $[0.8, 4]$. As for advertiser 2, there are three cases to consider:

- $q = \text{SF}$ and the refined scheme T is applied: realized value range is $[1.2, 2]$ and virtual value range is $[0.4, 2]$;

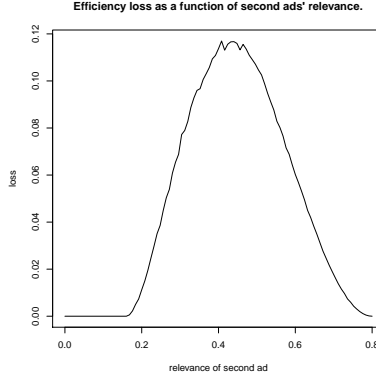


Figure 2: Efficiency loss of the revenue-optimal auction versus the efficiency-optimal auction for example in Section 3.2.5 as a function of the relevance of the second advertiser.

- $q = \neg \text{SF}$ and the refined scheme T is applied: realized value range is $[3\epsilon, 5\epsilon]$ and virtual value range is $[\epsilon, 5\epsilon]$;
- The coarse scheme \bar{T} is applied: realized value range is $[0.3, 0.5]$ and virtual value range is $[0.1, 0.5]$.

Applying the refined prediction scheme T which uses the location data lowers the expected welfare: Observe that the realized value of the advertiser 1 is always higher, and when \bar{T} is applied its virtual value is always higher as well, guaranteeing an efficient allocation. But because when $q = \text{SF}$ relevance predictions of the advertisers become closer, the range of advertiser 2's refined virtual value overlaps that of advertiser 1, so advertiser 2 sometimes wins despite this being inefficient.

How often would we expect such inefficiencies due to non flip-spread refinements in general? Assuming the setting of parameters as above, suppose we vary the second advertiser's relevance from 0 towards 0.8, and plot efficiency loss against the optimally efficient outcome; see Figure 3.2.5. As the figure shows, several refinements that are not flip-spread would still result in an efficiency increase. For instance, any refinement where $\bar{p}_2 > 0.4$ and p_2 is in the range $[\bar{p}_2, 0.8]$ would cause an efficiency increase. However, when $\bar{p}_2 < 0.4$ and p_2 is in the range $[\bar{p}_2, 0.4]$, the corresponding refinement causes an efficiency drop. Thus, we would expect that if the relevances of advertisers were initially roughly comparable (recall that for the first advertiser, $p_1 = \bar{p}_1 = 0.8$), any refinement ought to improve expected efficiency.

3.2.6 Non-i.i.d. advertisers

The previous example can be adjusted such that the resulting setting is completely equivalent, but now T is a flip-spread refinement of \bar{T} . This is by noticing that if advertisers are allowed to be non-i.i.d., a flip-spread refinement can make them more similar instead of more distinguished. For example, let advertiser 2's value be uniform over $[1.2, 2]$ instead of $[3, 5]$, and assume that finding out the user is in SF makes the advertisers' relevances flip from, say, 0.8, 0.25 to 0.8, 1. The ranges of their realized values however get closer, and the virtual value ranges overlap as above, leading to inefficiency.

4 Should the Search Engine Refine Predictions?

The previous section focused on data revelation that increases the welfare of all Pareto optimal auctions, and identified sufficient conditions under which signaling improves welfare. In this section we investigate the search engine’s incentive to perform refinements. We assume that the search engine optimizes a fixed convex combination of revenue and efficiency (see Section 2.4 for justification of this assumption), and show that refinement only improves the mixed objective. This generalizes [10], who show the same result for the revenue-optimal mechanism.

Theorem 4.1 (Refinement improves trade-off objective). *Consider a position auction and let M be a Pareto optimal mechanism. Let T be a prediction scheme that is a refinement of another prediction scheme \bar{T} . Then running M with T improves the objective of M in expectation in comparison to running M with \bar{T} .*

Proof. By Lemma 2.11, M should maximize the expected realized α -virtual surplus. For every fixed search query q , by Lemma 2.10, ranking the advertisers according to their realized α -virtual values maximizes the realized α -virtual surplus, and this is achieved by using T instead of \bar{T} . Taking expectation over q completes the proof. \square

The implication of the above theorem is that the seller should use as refined a prediction scheme as possible.

Notice that this result is less conditional than Theorem 3.1, that is the MHR, i.i.d. values and flip-spread refinement assumptions are not necessary. In fact, the fact that no condition of flip-spread is imposed on the refinement means that there must be a non-trivial trade-off between efficiency and revenue. In terms of efficiency, refining ad infinitum will not always be the right thing to do. But given that the search engine has already fixed a desired trade-off among efficiency and revenue, the best thing to do in terms of this trade-off is to use the α -Pareto optimal mechanism, and refine as far as possible.

Our main corollary follows immediately from Theorem 4.1 and Theorem 3.1, and states that prediction refinement can simultaneously increase the welfare and the expected objective of every Pareto optimal mechanism. In particular this holds for the Myerson mechanism.

Corollary 4.2 (Refinement improves welfare and trade-off). *Consider an i.i.d., MHR position auction and let M be a Pareto optimal mechanism. Then for every coarse prediction scheme \bar{T} and flip-spread refinement T , running M with T increases both welfare and M ’s objective in expectation, in comparison to running M with \bar{T} .*

5 Relation to Signaling

5.1 Our Result Applied to Signaling

In our model of relevance prediction, the features of the search query that determine the relevance scores can be viewed as private information held by the seller. This information, despite being unknown to the bidders, determines their realized values for the items. Auctions with seller information appear in the seminal work of Milgrom and Weber [21] and have been extensively studied in the economic literature (see, e.g., [3, 17]), and recently also in the computer science literature, where they’re sometimes called *probabilistic* auctions [8, 22, 10]. This reflects the fact that when the seller does not reveal his information, the bidders only

know its distribution and so the realized values are stochastic. Our model is a special case in which the seller information affects the realized values *multiplicatively*. We also assume that all privately held information (of the advertisers and seller) is independent.

The above previous works focus on *signaling schemes*, by which the seller communicates his information to the bidders. A signaling scheme maps the seller’s information to a possibly random output signal (a deterministic scheme is called a clustering scheme - see [11]). Based on the observed output signal, the bidders adjust their realized values according to their posterior belief about the seller’s information. Overloading notation, we denote the belief when no information is released by \bar{t} and after seeing the output signal by t . Observe that prediction refinement is mathematically equivalent to signaling, since given the features of the search query, the seller can either use a general query part \bar{t} to predict relevance, or use more information in form of a refined subpart t . In practice, prediction refinement avoids the communication costs involved in signaling information to the advertisers,¹³ and so may be desirable when many features are involved, provided that the seller knows how to translate the reported values into realized values.

We can now state our main result in the context of signaling, under the assumption that the seller’s information determines multiplicative factors by which bidder values are scaled. Let \bar{T} determine the multiplicative factors without revealing information and let T replace \bar{T} when information is revealed. We say that the signaling scheme is flip-spread if T is a flip-spread refinement of \bar{T} . Then a flip-spread signaling scheme for an i.i.d., MHR position auction and Pareto optimal mechanism improves both efficiency and the mechanism’s objective in expectation.

Levin and Milgrom [17] discuss how fine-grained signaling can affect *display* ad auctions. One of the topics they study is how revenue is impacted by ‘thin auctions’. Our results apply to their setting (under the assumption that signals are multiplicatively combined).

5.2 Relation to the Linkage Principle

The fundamental linkage principle of [21] states that full revelation of seller information is revenue-optimal (see also [14, Chapters 6 and 7]). In Appendix A.2 we highlight the differences of our work in terms of the model, mechanism and result statement.

6 Open Questions

Our results rely on the multiplicative assumption, which is natural in the context of position auctions. Can they be generalized to include additional effects that refinement may have on realized values, perhaps using some linear approximation? How can the notion of *flip-spread* refinement be generalized?

We have shown that for the Myerson mechanism, flip-spread refinements improves both welfare and revenue. For which other mechanisms does this desirable property hold?

¹³This is not the case in display ads, in which some engines report the user type to the advertisers and allow them to update their bids arbitrarily.

References

- [1] AGGARWAL, G., GOEL, A., AND MOTWANI, R. 2006. Truthful auctions for pricing search keywords. In *Proc. 8th ACM Conf. on Electronic Commerce (EC)*.
- [2] ALON, N. AND SPENCER, J. H. 2008. *The Probabilistic Method*. Wiley.
- [3] BOARD, S. 2009. Revealing information in auctions: the allocation effect. *Economic Theory* 38, 1, 125–135.
- [4] CLARKE, E. H. 1971. Multipart pricing of public goods. *Public Choice* 11, 17–33.
- [5] DIAKONIKOLAS, I., PAPADIMITRIOU, C., PIERRAKOS, G., AND SINGER, Y. 2012. Efficiency-revenue trade-offs in auctions. In *Proc. of ICALP*.
- [6] EDELMAN, B., OSTROVSKY, M., AND SCHWARZ, M. 2007. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review* 97, 1, 242–259.
- [7] EDELMAN, B. AND SCHWARZ, M. 2010. Optimal auction design and equilibrium selection in sponsored search auctions. *American Economic Review* 100, 2, 597–602.
- [8] EMEK, Y., FELDMAN, M., GAMZU, I., LEME, R. P., AND TENNENHOLTZ, M. 2012. Signaling schemes for revenue maximization. In *ACM Conference on Electronic Commerce*. 514–531.
- [9] EWERHART, C. 2009. Optimal design and ρ -concavity. Working Paper.
- [10] FU, H., JORDAN, P., MAHDIAN, M., NADAV, U., TALGAM-COHEN, I., AND VASSILVITSKII, S. 2012. Ad auctions with data. In *SAGT’12 Proceedings of the 5th international conference on Algorithmic Game Theory*. 168–179.
- [11] GHOSH, A., NAZERZADEH, H., AND SUNDARARAJAN, M. 2007. Computing optimal bundles for sponsored search. In *Workshop on Internet and Network Economics (WINE)*.
- [12] GRAEPEL, T., CANDELA, J. Q., BORCHERT, T., AND HERBRICH, R. 2010. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft’s bing search engine. In *Proceedings of the 27th International Conference on Machine Learning*.
- [13] GROVES, T. 1973. Incentives in teams. *Econometrica* 41, 617–631.
- [14] KRISHNA, V. 2010. *Auction Theory* Second Ed. Academic Press.
- [15] LAHAIE, S. AND PENNOCK, D. M. 2007. Revenue analysis of a family of ranking rules for keyword auctions. In *Proceedings of the 8th ACM conference on Electronic commerce*. EC ’07. ACM, New York, NY, USA, 50–56.
- [16] LAHAIE, S., PENNOCK, D. M., SABERI, A., AND VOHRA, R. V. 2007. Sponsored search auctions. In *Algorithmic Game Theory*, N. Nisan, T. Roughgarden, É. Tardos, and V. Vazirani, Eds. Cambridge University Press, Chapter 28, 699–716.

- [17] LEVIN, J. AND MILGROM, P. 2010. Online advertising: Heterogeneity and conflation in market design. *American Economic Review: Papers and Proceedings* 100, 603–607.
- [18] LIKHODEDOV, A. AND SANDHOLM, T. 2003. Auction mechanism for optimally trading off revenue and efficiency. In *EC '03: Proceedings of the 4th ACM conference on Electronic commerce*. ACM, New York, NY, USA, 212–213.
- [19] MCAFEE, R. AND MCMILLAN, J. 1987. Auctions and bidding. *Journal of Economic Literature* 25, 2, 699–738.
- [20] MEHTA, A., SABERI, A., VAZIRANI, U., AND VAZIRANI, V. 2002. Adwords and generalized on-line matching. In *Proc. 46th IEEE Symp. on Foundations of Computer Science (FOCS)*.
- [21] MILGROM, P. AND WEBER, R. J. 1982. A theory of auctions and competitive bidding. *Econometrica* 50, 1089–1122.
- [22] MILTERSEN, P. B. AND SHEFFET, O. 2012. Send mixed signals: earn more, work less. In *ACM Conference on Electronic Commerce*. 234–247.
- [23] MYERSON, R. 1981. Optimal auction design. *Mathematics of Operations Research* 6, 1, 58–73.
- [24] MYERSON, R. AND SATTERTHWAIT, M. 1983. Efficient mechanisms for bilateral trade. *Journal of Economic Theory* 29, 1, 265–281.
- [25] PERRY, M. AND RENY, P. 1999. On the failure of the linkage principle in multi-object auctions. *Econometrica* 67, 4, 885–890.
- [26] VARIAN, H. R. 2007. Position auctions. *International Journal of Industrial Organization* 25, 6, 1163–1178.
- [27] VICKREY, W. 1961. Counterspeculation, auctions, and competitive sealed tenders. *J. of Finance* 16, 8–37.

A Appendix

A.1 Analysis of Example 3.2 (Efficiency Loss Due to Refinement for Regular Distribution)

In this appendix we provide the calculations for Example 3.2, which demonstrates that when advertisers’ i.i.d. values are drawn from a regular but non-MHR distribution, the expected welfare of Myerson can decrease following a flip-spread refinement.

Define Δ to be the difference between the expected efficiency loss due to refinement and the expected efficiency loss due to coarseness, when the values are drawn from the truncated shifted equal revenue distribution defined above in the example. In this analysis we use the notation s instead of v . Recall that \bar{s} is such that $\varphi(\bar{s}) = 2\varphi(s')$.

A.1.1 Step 1

In this step we show that

$$\Delta = \frac{1}{2} \int_2^{H+1} \left((\bar{s} - 2s')F(\bar{s}) + s'F(s') - \int_{s'}^{\bar{s}} F(s)ds \right) f(s')ds'.$$

Consider the difference for fixed s' :

$$\begin{aligned} \int_{s'}^{\frac{s'}{c}} (s' - cs)f(s)ds + \int_{\frac{s'}{c}}^{\bar{s}} (s' - cs)f(s)ds &= \int_{s'}^{\bar{s}} (s' - cs)f(s)ds \\ &= s' \int_{s'}^{\bar{s}} f(s)ds - c \int_{s'}^{\bar{s}} sf(s)ds. \end{aligned}$$

We integrate by parts:

$$\begin{aligned} \int_{s'}^{\bar{s}} f(s)sds &= F(s)s \Big|_{s'}^{\bar{s}} - \int_{s'}^{\bar{s}} F(s)ds \\ &= \bar{s}F(\bar{s}) - s'F(s') - \int_{s'}^{\bar{s}} F(s)ds. \end{aligned}$$

Plugging in:

$$\begin{aligned} c\bar{s}F(\bar{s}) - cs'F(s') - c \int_{s'}^{\bar{s}} F(s)ds - s'F(\bar{s}) + s'F(s') &= \\ (c\bar{s} - s')F(\bar{s}) + (s' - cs')F(s') - c \int_{s'}^{\bar{s}} F(s)ds &= \\ c \left((\bar{s} - \frac{s'}{c})F(\bar{s}) + (\frac{s'}{c} - s')F(s') - \int_{s'}^{\bar{s}} F(s)ds \right) \end{aligned}$$

A.1.2 Finding \bar{v}

$$\begin{aligned} \frac{(\bar{s} - 1)^2 + H}{H} &= \frac{2(s' - 1)^2 + 2H}{H} \\ \bar{s} &= 1 + \sqrt{2(s' - 1)^2 + H} \end{aligned}$$

A.1.3 The expressions to integrate

We use the following:

$$\begin{aligned} \int_{s'}^{\bar{s}} F(s)ds &= \frac{H}{H-1} \int_{s'}^{\bar{s}} \left(1 - \frac{1}{s-1}\right)ds \\ &= \frac{H}{H-1} \left(s - \log(s-1) \Big|_{s'}^{\bar{s}} \right) \\ &= \frac{H}{H-1} \left(\bar{s} - \log(\bar{s}-1) - s' + \log(s'-1) \right) \end{aligned}$$

So the inner function before multiplying by $f(s')$ is:

$$\begin{aligned}
& \frac{1}{2} \left((\bar{s} - 2s')F(\bar{s}) + s'F(s') - \int_{s'}^{\bar{s}} F(s)ds \right) = \\
& \frac{H}{2(H-1)} \left((\bar{s} - 2s')\left(1 - \frac{1}{\bar{s}-1}\right) + s'\left(1 - \frac{1}{s'-1}\right) - \bar{s} + \log(\bar{s}-1) + s' - \log(s'-1) \right) = \\
& \frac{H}{2(H-1)} \left(\bar{s} - 2s' - \frac{\bar{s}-2s'}{\bar{s}-1} + s' - \frac{s'}{s'-1} - \bar{s} + \log(\bar{s}-1) + s' - \log(s'-1) \right) = \\
& \frac{H}{2(H-1)} \left(\log(\bar{s}-1) - \frac{\bar{s}-2s'}{\bar{s}-1} - \frac{s'}{s'-1} - \log(s'-1) \right)
\end{aligned}$$

Multiplying by $f(s') = \frac{H}{H-1} \cdot \frac{1}{(s'-1)^2}$ and plugging in $\bar{s} = 1 + \sqrt{2(s'-1)^2 + H}$, we get three integrals as follows (we write them without the factor of $\frac{1}{2} \left(\frac{H}{H-1} \right)^2$ for simplicity):

$$\begin{aligned}
\frac{1}{2} \int_{1-b}^{H-b} \frac{\log(2(s'-1)^2 + H)}{(s'-1)^2} ds' &= \frac{1}{2} \int_1^H \frac{\log(2x^2 + H)}{x^2} dx \\
\int_{1-b}^{H-b} \frac{1 + \sqrt{2(s'-1)^2 + H} - 2s'}{(s'-1)^2 \sqrt{2(s'-1)^2 + H}} ds' &= \int_1^H \frac{\sqrt{2x^2 + H} - 2x - 1}{x^2 \sqrt{2x^2 + H}} dx \\
&= \int_1^H x^{-2} dx - \int_1^H \frac{2x+1}{x^2 \sqrt{2x^2 + H}} dx \\
\int_{1-b}^{H-b} \frac{s'}{(s'-1)^3} + \frac{\log(s'-1)}{(s'-1)^2} ds' &= \int_1^H \frac{x+1}{x^3} + \frac{\log x}{x^2} dx
\end{aligned}$$

A.1.4 Calculating the integrals

First integral:

$$\begin{aligned}
& \frac{\sqrt{2} \arctan(\frac{\sqrt{2}x}{\sqrt{H}})}{\sqrt{H}} - \frac{\log(H + 2x^2)}{2x} \Big|_1^H = \\
& \frac{\sqrt{2} \arctan(\frac{\sqrt{2}H}{\sqrt{H}})}{\sqrt{H}} - \frac{\log(H + 2H^2)}{2H} - \frac{\sqrt{2} \arctan(\frac{\sqrt{2}}{\sqrt{H}})}{\sqrt{H}} + \frac{\log(H + 2)}{2} = \\
& \frac{\sqrt{8H} \left(\arctan(\sqrt{2H}) - \arctan(\sqrt{2/H}) \right) - \log H - \log(1 + 2H) + H \log(H + 2)}{2H} = 3.51487
\end{aligned}$$

Second integral:

$$1 - \frac{1}{H} - \left(-\frac{\sqrt{H + 2x^2}}{Hx} - \frac{2 \log(\sqrt{H(H + 2x^2)} + H)}{\sqrt{H}} + \frac{2 \log(\sqrt{H}x)}{\sqrt{H}} \Big|_1^H \right) = 0.7298$$

Third integral:

$$\begin{aligned}
-\frac{4x + 2x \log x + 1}{2x^2} \Big|_1^H &= \frac{5}{2} - \frac{4H + 2H \log H + 1}{2H^2} \\
&= 2.4911
\end{aligned}$$

A.1.5 Putting everything together

$$\frac{1}{2} \left(\frac{H}{H-1} \right)^2 (3.51487 - 0.7298 - 2.4911) = 0.1473.$$

The final answer is positive, and so we have shown that the inefficiencies due to refinement surpass those due to coarseness in this example.

A.2 Relation to the Linkage Principle

The Milgrom-Weber model is more general than ours in that realized values depend on the seller's information in an arbitrary way (and can also depend on other bidders' private values). This dependence however must be the same for all bidders. This symmetry requirement is crucial, and without it releasing seller information may actually harm revenue [25, 14, Chapter 8]. In contrast, an inherent feature of our model is that advertisers can be asymmetric or even antisymmetric in the way their relevance changes as more features are used for prediction. We remark that the Milgrom-Weber model also allows a certain form of correlation among private information called *affiliation* (see [14, Appendix D] and [2, FKG inequality]).

Milgrom and Weber do not consider direct revelation mechanisms as we do, since in their model the information of the seller and other bidders can affect realized values arbitrarily. Instead they analyze first price, second price and English auctions assuming symmetric (possibly untruthful) equilibrium bidding. Their result that using full seller information is optimal holds even for the second price auction with no reserve. This is not the case in our model, where we show that the combined objective of a Pareto optimal mechanism increases given refined prediction.